The NIST Chemistry WebBook: A Chemical Data Resource on the Internet

P. J. Linstrom and W. G. Mallard

National Institute of Standards and Technology

Physical and Chemical Properties Division

100 Bureau Drive, Stop 8380

Gaithersburg, MD 20899-8380

The NIST Chemistry WebBook (http://webbook.nist.gov) is an Internet site that provides access to chemical and physical property data both from NIST and other sources. The site was established in 1996 and has grown to encompass a wide variety of thermochemical, ion energetics, solubility and spectroscopic data. Thermochemical data available includes heats of formation, heats of phase transitions, and heat capacities. Thermochemical properties of many reactions that support heat of formation values are provided. Automated tools are used to check data prior to its inclusion in the web site. Most of the collections in the site provide extensive coverage of the literature in their field and include relevant metadata such as the experiment type or important auxiliary data. These features make the site an excellent tool for data evaluation.

A major goal of the project was to provide convenient access to any types of chemical data. Several challenges were encountered in he development of systems and conventions for concisely and accurately displaying chemical data on the Internet. The next phase of the evolution of the site will be the addition of tools to id researchers in getting data from the site.

Data from the site has found applications in industrial, research and educational settings. Usage patterns for the site will be discussed.

Introduction

The NIST Chemistry WebBook (http://webbook.nist.gov) is an Internet site that provides physical and chemical reference data to users on the World Wide Web. The site provides a wide variety of data that can be used by many disciplines. The site is part of the NIST Standard Reference Data Program, a Congressionally mandated program[1] responsible for making critically evaluated reference data readily available to scientists, engineers and the general public.

Data Available

The data available in the site is composed of collections that were either compiled at NIST or contributed to NIST. The major data collections in the site are summarized in table 1.

Data available in the site includes collections which were previously published elsewhere[2,3,4,5], and updates of collections previously published elsewhere[6,7].

Data is organized by chemical species and broken down into 12 different groups. The data groups allow control of the data that is displayed on a page. Four of the groups refer to neutral thermochemical data types, two refer to ion energetics data types, one refers to solubility data, and the remaining groups refer to spectroscopic data.

Thermochemical data available at the site for individual species includes enthalpies of formation, entropies, constant pressure heat capacities, critical temperatures and pressures, enthalpies of phase transitions and vapor pressures (Antoine equation parameters). Data provided for reactions includes enthalpies, entropies and Gibbs free energy values. As shown in table 2, the amount of data currently available varies widely by property. Since the site contains multiple measurements for many properties, the number of species and reactions with data is generally less than the total number of data points.

Updates to the site are carried out periodically. The most recent update occurred in February 2000, and is the sixth release of the site.

In addition to the data described above the site contains a facility that contains models for 16 common fluids. These models provide thermophysical properties for user specified conditions. Data may be found for points along isotherms, isobars, and the saturation curve.

Quality of Data

The data in the site is well documented and compiled by researchers familiar with the data and the literature it comes from. With few exceptions, all of the data in the site comes from the peer-reviewed literature. In many of the collections contained in the site, useful auxiliary data such as the experimental method used and relevant comments are included along with the data. Many of the collections attempt to provide comprehensive coverage of the literature and thus contain multiple measurements of the same property.

Data complied at NIST is subject to additional quality control procedures. For example the neutral thermochemical data compiled by Afeefy, Liebman and Stein is processed to detect outliers. One test conducted is to look for large standard deviations in multiple determinations of a value. Cases with large standard deviations are examined for possible errors. This examination has been done for both gas and condensed phase enthalpies of formation.

Regardless of the source of the data, care is taken to preserve data as it is recorded. Data is stored as text in the units in which it was recorded. Data compiled at NIST is usually recorded in the units used in the original publication to reduce the chance of errors. The software for the site performs any needed unit conversions and preserves the appropriate number of significant digits in the result.

Because of the quality and variety of data available, the site can be a good tool for evaluation of thermochemical data. The site often provides access to related measurements and bibliographic information. For example, evaluation of gas phase enthalpy of formation data is aided by access to supporting heat of reaction data, condensed phase enthalpy of formation data, and heat of phase change data.

The site contains a feature that allows users to submit error reports. These reports are logged to a database at NIST and forwarded to the maintainers of the relevant collections.

Data Presentation

The site was designed so that data would be clear and easy to find regardless of the user's background and browser software. Provisions have been made to support both older and newer web browsers. The site is able to display Greek letters and other special characters found in chemical data notation on a wide range of browsers.

Since data is organized by chemical species, users simply need to search for the species of interest to find the data they need. The site provides several search options for looking up chemical species, including searches based on the chemical formula, name, CAS registry number or structure of the species. Several searches based on physical quantities are also available.

All data pages include information identifying the species followed by one or more groups of data. Identifying information is displayed at the top portion of the page as shown in figure 1. This information includes the species name, formula, and molecular weight. When available, the CAS registry number, a structure drawing, and additional names are also displayed. The top portion of the page also includes links to data for the species.

Most data in the site is presented in tabular form as illustrated in figure 2. Tables in the site were designed to be clear and readable when printed. References in the tables are linked to a bibliography at the bottom of the page and method codes are linked to pages that explain the codes. The references in the bibliography contain links that search for other articles by a given author and for other species that contain data from the source. The collection(s) from which the data are drawn are identified at the top of the section containing the data. This information contains links to pages that describe the recommended citation for the data.

Where appropriate, Java applets have been made available to allow users to view data graphically. These applets allow users to interactively re-scale graphs of infrared and mass spectra and selected curves of thermophysical data. This enables users to examine the data more closely than would be possible if the data were presented as a static image. Static images of spectra are provided for users that cannot display the applets.

When published in printed form, many of the data sets in the site contained introductory material describing the data and the manner in which they were collected. This information

can be of great use to users, and where possible was included as part of the documentation for the web site.

Usage Patterns

There are currently a significant number of users using the site and there is evidence that many users return on a regular basis. A summary of usage over the life of the site is presented in figure 3. The data in figure 3 excludes requests for graphic images, class files, and other requests that are not associated with loading a new page. During the transition from the second release to the third release of the site (August 1997), the format of the data pages changed. This change appears to have lead to new usage patterns resulting in a significant increase in pages viewed. It should be noted that the large usage numbers observed for November 1999 are due to a script that was run against the site at the time. The bar on figure 3 has been truncated for this month; the actual value for the bar should be 1,170,000 pages.

Figure 4 shows the percentage of Internet addresses that have visited the site in a week which had visited the site in a prior week. The data in the figure covers the third, fourth and fifth releases of the site (August 19, 1997 to February 6, 2000). Since the server can only log Internet addresses rather than individual users, different users who share the same Internet service provider or corporate proxy server will show up as the same user. This means that the number of users is greater than the number of Internet addresses that access the site and that different users with the same proxy server could show up as a returning user. Because of this the actual return rate for users is probably slightly less than that shown

in figure 4. During the period covered by the figure, the site had, on average, users from 7000 different Internet addresses per week.

The top level internet domains which accessed the site most often during the fifth release are summarized in table 3. This table was generated by looking up names for Internet addresses that accessed the site during this time. Approximately one quarter of the Internet addresses lacked a DNS entry and thus could not be looked up. The percentage figures given in table 3 were normalized to exclude usage from addresses that could not be looked up. Addresses from total of 146 different top level domains accessed the site during this period. At least one fifth of the .com and .net addresses are known Internet service providers. E-mail messages from users have indicated that the site is used by both large and small companies and educational institutions ranging from elementary schools to universities.

Although the log files indicate the Internet addresses of users of the site, additional information such as the species searched for and the data examined is not logged. This is done to protect the privacy of users.

Future Trends

The NIST Chemistry WebBook is an ongoing project. Work continues on adding new data and features to the site. As always, NIST welcomes contributions of high quality documented data sets that are appropriate for the site. Two major enhancements are planned

for the site: the addition of selected data from the Thermodynamics Research Center Source Database[8] and the addition of additional data export options.

The new data export options will allow data to be exported in a manner that can be used by computer software. This will be done to make it easier to enable users to use data from the site in research problems. One major hurdle to overcome in the development of this feature is the selection of an appropriate format for exporting data.

Conclusions

The NIST Chemistry WebBook has succeeded in providing convenient access to a wide range of physico-chemical data. The overwhelming majority of the data comes from the peer-reviewed literature and is well documented with auxiliary data where appropriate. Steps are taken to insure the quality of the data available in the site and correct mistakes that do occur.

A significant and growing number of people are using the site. The large percentage of returning users indicates that many users find the information at the site of use. Analysis of log files indicates that the site is used by people worldwide, with significant usage by domestic educational institutions, commercial organizations, and users connecting through internet service providers.

Plans are underway to continue the evolution of the site, so that it will continue to serve its mission to deliver quality physical and chemical property data to engineers, researchers and educators.

Acknowledgement

The NIST Chemistry WebBook was developed and is maintained in part with support from the NIST Systems Integration for Manufacturing Applications (SIMA) program.

Literature Cited

(1) Standard Reference Data Act, United States Code, Jul 1968. U.S. Pub. L. 90-396, 82 Stat. 339.

(2) Chase, Jr., M. W. NIST-JANAF thermochemical tables (fourth edition). *J. Phys. Chem. Ref. Data, Monograph 9* **1998**, 1-1963.

(3) Huber, K. P.; Herzberg, G. *Molecular Spectra and Molecular Structure. IV. Constants of Diatomic Molecules.* Van Nostrand Reinhold Co.: New York, 1979.

(4) Shimanouchi., T. *Tables of Molecular Vibrational Frequencies, Consolidated Volume 1*. Nat. Bur. Stand. (U.S.): Washington, 1972.

(5) Hunter, E. P. L.; Lias, S.G. Evaluated gas phase basicities and proton affinities of molecules: An update. *J. Phys. Chem. Ref. Data*, **1998**, 27, 413-656.

(6) Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. Gas-phase ion and neutral thermochemistry. *J. Phys. Chem. Ref. Data, Suppl. 1* **1988**, 17, 1-861.

(7) Jacox, M. E. *Vibrational and electronic energy levels of polyatomic transient molecules. Supplement A*. J. Phys. Chem. Ref. Data, **1998**, 27, 115-393.

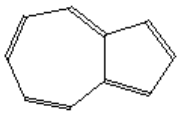
(8) Themodynamics Research Center *TRC source database v. 1.02 for Windows 95*, 98, NT Texas A&M University: College Station, TX, 2000.

Figures

NIST Standard Reference Data Program | Online Databases | Chemistry WebBook

Azulene

- **Formula:** $C_{10}H_8$
- **Molecular Weight:** 128.17
- **CAS Registry Number:** 275-51-4
- **Chemical Structure:**



This structure is also available as a2d Mol file or as a computed 3d Mol file.

- **Other Names:** Bicyclo[5.3.0]decapentaene; Cyclopentacycloheptene; Azunamic; Bicyclo (5.3.0)-1,3,5,7,9-decapentaene; Bicyclo(0.3.5)deca-1,3,5,7,9-pentaene; BICYCLO(5.3.0)-DECA-2,4,6,8,10-PENTAENE
- Notes / Error Report
- **Other Data Available:**
  - Gas phase thermochemistry data
  - Condensed phase thermochemistry data
  - Phase change data
  - Reaction thermochemistry data
  - Gas phase ion energetics data
  - Ion clustering data
  - Gas Phase IR Spectrum
  - Mass Spectrum
  - UV/Visible Spectrum
- Switch to calorie-based units

Figure 1: Top portion of data page from the NIST Chemistry WebBook.

Gas phase thermochemistry data

Go To: Top, References, Notes / Error Report

Data compiled as indicated in comments:
ALS - H.Y. Afeefy, J.F. Liebman, and S.E. Stein
GT - Glushko Thermocenter, Russian Academy of Sciences, V.S. Yungman, director

Quantity	Value	Units	Method	Reference	Comment
$\Delta_f H°_{gas}$	308.	kJ/mol	Chyd	Roth, Bohm, et al., 1983	*ALS*
$\Delta_f H°_{gas}$	280.	kJ/mol	Ccb	Kovats, Gunthard, et al., 1957	Correction to Kovats, Gunthard, et al., 1955; *ALS*

Constant pressure heat capacity of gas.

$C_{p,gas}$ (J/mol*K)	Temperature (K)	Reference	Comment
81.96	200.	Kovats E., 1955	*GT*
128.41	298.15		
129.41	300.		
176.36	400.		
216.27	500.		
248.19	600.		
274.30	700.		
295.35	800.		
312.75	900.		
327.36	1000.		

Figure 2: Part of a data page for azulene from the NIST Chemistry WebBook.

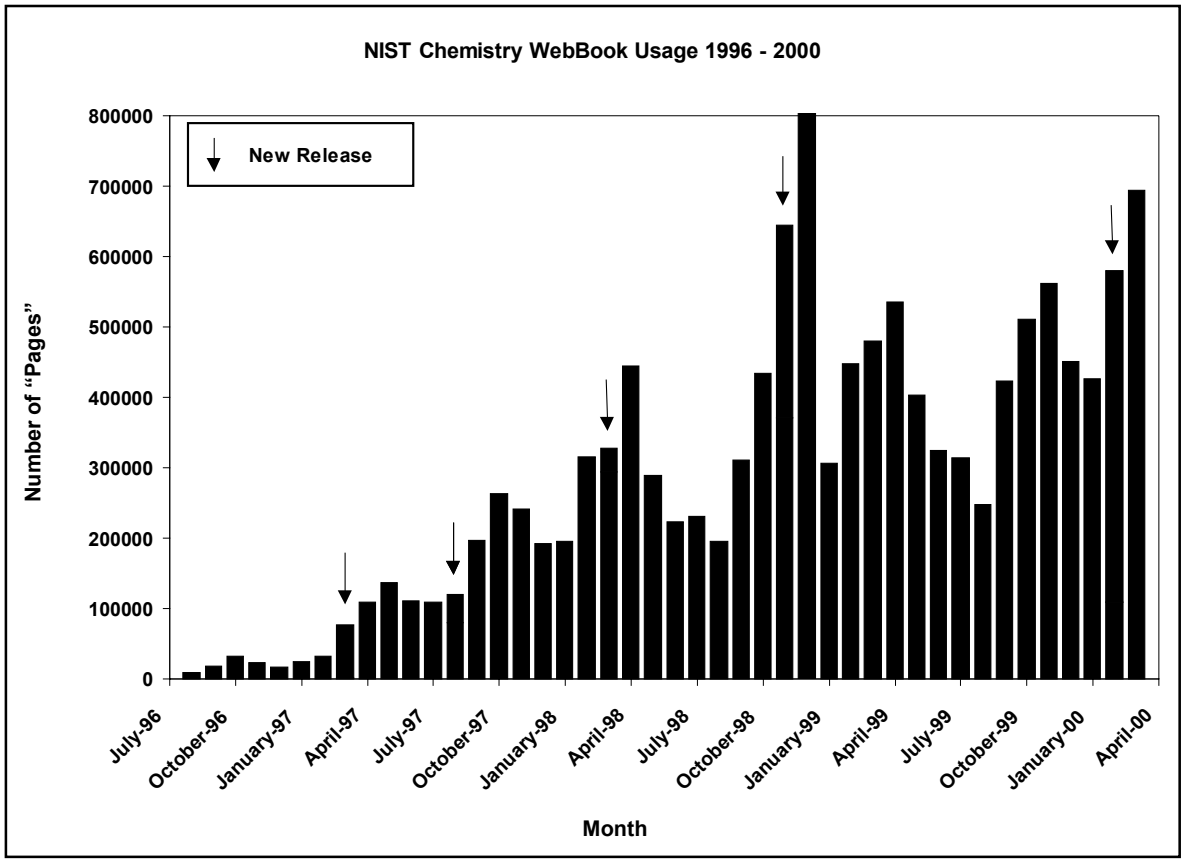


Figure 3: NIST Chemistry WebBook usage August 1996 to April 2000.

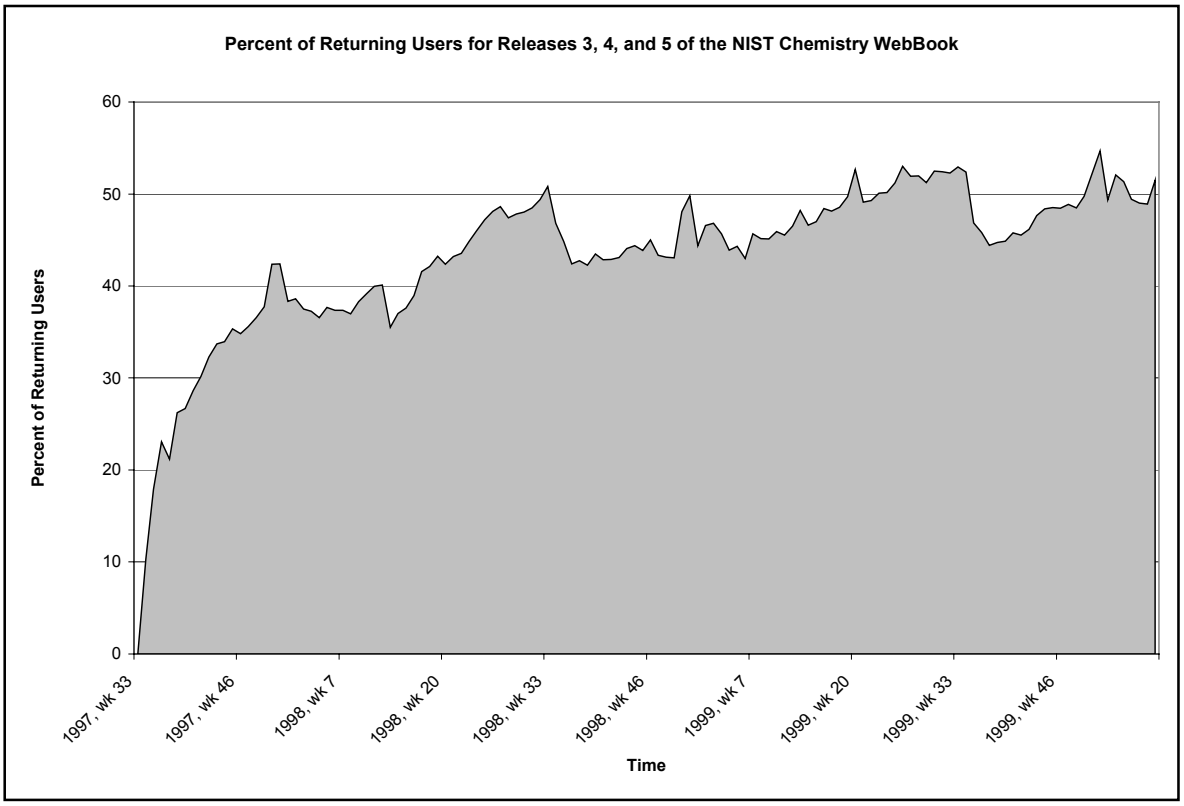


Figure 4: Percent of returning users for releases 3, 4, and 5 of the NIST Chemistry WebBook.

Tables

Data Type	Compiler(s)
Neutral thermochemical data	H.Y. Afeefy, J.F. Liebman, S.E. Stein
Condensed phase heat capacity data	E.S. Domalski, E.D. Hearing
Positive ion energetics data	S.G. Lias, H.M. Rosenstock, K. Draxl, B.W. Steiner, J.T. Herron, J.L. Holmes, R.D. Levin, J.F. Liebman, S.A. Kafafi
Negative ion energetics data	J.E. Bartmess
Cluster ion thermochemistry data	M.M. Meot-Ner(Mautner)
Proton affinity data	E.P. Hunter, S.G. Lias
Vibrational and electronic energy level data	M.E. Jacox
Vibrational Frequency data	T. Shimanouchi
Constants of diatomic molecules	K.P. Huber, G. Herzberg
Organometallic thermochemical data	J.A. Martinho Simões
Heat of sublimation data	J.S. Chickos
Boiling point data	R.L. Brown, S.E. Stein
Henry's Law Constants	R. Sander
Thermophysical properties of fluids	E.W. Lemmon, M.O. McLinden, D.G. Friend
UV/Visible Spectra	V. Talrose, E.B. Stern, A.A. Goncharova, N.A. Messineva, N.V. Trusova, M.V. Efimkina
Names, 2-D structures, mass and IR spectra data	NIST Mass Spectrometry Data Center, S.E. Stein, Director
3-D structures	K.K. Irikura

Table 1: Collections of data contained in the February 2000 release of the NIST Chemistry WebBook.

Property	Total Data Points
$\Delta_f H°$, gas	7452
$S°$, gas	1106
C_p , gas	541
$\Delta_f H°$, liq	2727
$S°$, liq	949
C_p, liq	2564
$\Delta_f H°$, cr	3203
$S°$, cr	1064
C_p, cr	1586
ΔH, phase change	7856
ΔS, phase change	1130
Heat of combustion	6530
Boiling point	9271
Melting point	365
Critical temperature	399
Critical pressure	153
Antoine parameters	1481
ΔH, reaction	10065
ΔS, reaction	6055
ΔG, reaction	722

Table 2: Total number of data points for selected thermochemical properties.

Percent of Total Bytes	Domain	
	Suffix	Meaning
28.14	.edu	USA Educational
19.06	.com	Commercial, mainly USA
12.51	.net	Network
5.05	.ca	Canada
4.85	.de	Germany
2.99	.jp	Japan
2.76	.fr	France
2.72	.gov	USA Government
2.25	.uk	United Kingdom
1.61	.nl	Netherlands
1.30	.ch	Switzerland
1.29	.au	Australia
1.18	.it	Italy
0.91	.se	Sweden
0.84	.es	Spain
0.72	.fi	Finland
0.68	.dk	Denmark
0.67	.us	United States
0.64	.tw	Taiwan
0.64	.mil	USA Military

Table 3: Usage by domain for the 5th release of the NIST Chemistry WebBook (November 1998 to February 2000).